

SEQUENCE LISTING

1) GENERAL INFORMATION:

(i) APPLICANT:

(A) NAME: Medical Research Council
(B) STREET: 20 Park Crescent
(C) CITY: London
(E) COUNTRY: United Kingdom
(F) POSTAL CODE (ZIP): W1N 4AL

(ii) TITLE OF INVENTION: Improvements in or Relating to Binding
Proteins for Recognition of DNA

(iii) NUMBER OF SEQUENCES: 18

(iv) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0. Version #1.30 (EPO)

:2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 60 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

CTCCTGCAGT TGGACCTGTG CCATGGCCGG CTGGGCCGCA TAGAATGGAA CAACTAAAGC

60

:2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 92 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

Met	Ala	Glu	Glu	Arg	Pro	Tyr	Ala	Cys	Pro	Val	Glu	Ser	Cys	Asp	Arg
1				5				10						15	
Arg	Phe	Ser	Arg	Ser	Asp	Glu	Leu	Thr	Arg	His	Ile	Arg	Ile	His	Thr
			20				25						30		

Gly Gln Lys Pro Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser Xaa
 35 40 45
 Xaa Xaa Xaa Leu Xaa Xaa His Xaa Arg Thr His Thr Gly Glu Lys Pro
 50 55 60
 Phe Ala Cys Asp Ile Cys Gly Arg Lys Phe Ala Arg Ser Asp Glu Arg
 65 70 75 80
 Lys Arg His Thr Lys Ile His Leu Arg Gln Lys Asp
 85 90

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 26 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

TATGACTTGG ATGGGAGACC GCCTGG

26

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 28 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

AATTCAGGC GGTCTCCCAT CCAAGTCA

28

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

TATATAGCGT GGGCGTATAT A

21

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs

264090" 80466/80

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

GCGTATATAC GCCCACGCTA TATA

24

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

TATATAGCGN NNGCGTATAT A

21

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

GCGTATATAC GCNNNGCTA TATA

24

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 33 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

TTCCAATGGAG ACGCAGAAGC CCTTCAGCGG CCA

33

(2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 33 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

08793408.060397

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

TTCCATGGAG ACGCAGGTGA GTTCCTCACG CCA

33

(2) INFORMATION FOR SEQ ID NO: 11:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 33 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

CCCCTTTCTC TTCCAGAAGC CCTTCAGCGG CCA

33

(2) INFORMATION FOR SEQ ID NO: 12:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 33 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS:

(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

Met Ala Glu Glu Lys Pro Phe Gln Cys Arg Ile Cys Met Arg Asn Phe
1 5 10 15

Ser Asp Arg Ser Ser Leu Thr Arg His Thr Arg His Thr Gly Glu Lys
20 25 30

Pro

(2) INFORMATION FOR SEQ ID NO: 13:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 33 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS:

(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

Met Ala Glu Glu Lys Pro Phe Gln Cys Arg Ile Cys Met Arg Asn Phe
1 5 10 15

08793408 060397

Ser Glu Arg Gly Thr Leu Ala Arg His Glu Lys His Thr Gly Glu Lys
 20 25 30

Pro

(2) INFORMATION FOR SEQ ID NO: 14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser Gln Gly Gly Asn Leu
 1 5 10 15

Val Arg His Leu Arg His Thr Gly Glu Lys Pro
 20 25

(2) INFORMATION FOR SEQ ID NO: 15:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 26 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser Gln Ala Gln Thr Leu
 1 5 10 15

Gln Arg His Leu Lys His Thr Gly Glu Lys
 20 25

(2) INFORMATION FOR SEQ ID NO: 16:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 26 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

08793408.060397

Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser Gln Ala Ala Thr Leu
1 5 10 15

Gln Arg His Leu Lys His Thr Gly Glu Lys
20 25

(2) INFORMATION FOR SEQ ID NO: 17:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 26 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser Gln Ala Gln Asp Leu
1 5 10 15

Gln Arg His Leu Lys His Thr Gly Glu Lys
20 25

(2) INFORMATION FOR SEQ ID NO: 18:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 89 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

Met Ala Glu Glu Lys Pro Phe Gln Cys Arg Ile Cys Met Arg Asn Phe
1 5 10 15

Ser Asp Arg Ser Ser Leu Thr Arg His Thr Arg Thr His Thr Gly Glu
20 25 30

Lys Pro Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser Asp Arg Ser
35 40 45

His Leu Thr Arg His Thr Arg Thr His Thr Gly Glu Lys Pro Phe Gln
50 55 60

Cys Arg Ile Cys Met Arg Asn Phe Ser Asp Arg Ser Asn Leu Thr Arg
65 70 75 80

His Thr Arg Thr His Thr Gly Glu Lys
85

08793408.060397

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- (i) APPLICANT: Choo, Yen
Klug, Aaron
Sanchez Garcia, Isidro
- (ii) TITLE OF INVENTION: Improvements in or Relating to
Binding Proteins for Recognition of DNA
- (iii) NUMBER OF SEQUENCES: 18
- (iv) CORRESPONDENCE ADDRESS:
 - (A) ADDRESSEE: Pillsbury Madison & Sutro, L.L.P.
 - (B) STREET: 1100 New York Avenue, N.W.
 - (C) CITY: Washington
 - (D) STATE: D.C.
 - (E) COUNTRY: USA
 - (F) ZIP: 20005-3918
- (v) COMPUTER READABLE FORM:
 - (A) MEDIUM TYPE: Floppy disk
 - (B) COMPUTER: IBM PC compatible
 - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
 - (D) SOFTWARE: Word Perfect
- (vi) CURRENT APPLICATION DATA:
 - (A) APPLICATION NUMBER:
 - (B) FILING DATE:
 - (C) CLASSIFICATION:
- (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: PCT/GB95/01949
 - (B) FILING DATE: 17-AUG-1995
- (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: GB 9514698.1
 - (B) FILING DATE: 18-JUL-1995
- (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: GB 9422534.9
 - (B) FILING DATE: 08-NOV-1994
- (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: GB 9416880.4
 - (B) FILING DATE: 20-AUG-1994

(2) INFORMATION FOR SEQ ID NO: 1:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 60 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

CTCCTGCAGT TGGACCTGTG CCATGGCCGG CTGGGCCGCA TAGAATGGAA 50
CAACTAAAGC 60

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 92 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

Met Ala Glu Glu Arg Pro Tyr Ala Cys Pro
 5 10
Val Glu Ser Cys Asp Arg Arg Phe Ser Arg
 15 20
Ser Asp Glu Leu Thr Arg His Ile Arg Ile
 25 30
His Thr Gly Gln Lys Pro Phe Gln Cys Arg
 35 40
Ile Cys Met Arg Asn Phe Ser Xaa Xaa Xaa
 45 50
Xaa Leu Xaa Xaa His Xaa Arg Thr His Thr
 55 60
Gly Glu Lys Pro Phe Ala Cys Asp Ile Cys
 65 70
Gly Arg Lys Phe Ala Arg Ser Asp Glu Arg
 75 80
Lys Arg His Thr Lys Ile His Leu Arg Gln
 85 90
Lys Asp

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 26 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

TATGACTTGG ATGGGAGACC GCCTGG

26

(2) INFORMATION FOR SEQ ID NO: 4:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 28 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

AATTCCAGGC GGTCTCCCAT CCAAGTCA

28

(2) INFORMATION FOR SEQ ID NO: 5:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 21 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

TATATAGCGT GGGCGTATAT A

21

(2) INFORMATION FOR SEQ ID NO: 6:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 24 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

GCGTATATAC GCCCAGCTA TATA

24

(2) INFORMATION FOR SEQ ID NO: 7:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 21 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

TATATAGCGN NNGCGTATAT A

21

(2) INFORMATION FOR SEQ ID NO: 8:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 24 base pairs

- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

GCGTATATAC GCNNNCGCTA TATA

24

(2) INFORMATION FOR SEQ ID NO: 9:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 33 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

TTCCATGGAG ACGCAGAAGC CCTTCAGCGG CCA

33

(2) INFORMATION FOR SEQ ID NO: 10:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 33 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

TTCCATGGAG ACGCAGGTGA GTTCCTCACG CCA

33

(2) INFORMATION FOR SEQ ID NO: 11:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 33 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

CCCCTTTCTC TTCCAGAAGC CCTTCAGCGG CCA

33

(2) INFORMATION FOR SEQ ID NO: 12:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 33 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

Met Ala Glu Glu Lys Pro Phe Gln Cys Arg
5 10

Ile Cys Met Arg Asn Phe Ser Asp Arg Ser
15 20

Ser Leu Thr Arg His Thr Arg His Thr Gly
25 30

Glu Lys Pro

(2) INFORMATION FOR SEQ ID NO: 13:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 33 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

Met Ala Glu Glu Lys Pro Phe Gln Cys Arg
5 10

Ile Cys Met Arg Asn Phe Ser Glu Arg Gly
15 20

Thr Leu Ala Arg His Glu Lys His Thr Gly
25 30

Glu Lys Pro

(2) INFORMATION FOR SEQ ID NO: 14:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 27 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

Phe Gln Cys Arg Ile Cys Met Arg Asn Phe
5 10

Ser Gln Gly Gly Asn Leu Val Arg His Leu
15 20

Arg His Thr Gly Glu Lys Pro
25

(2) INFORMATION FOR SEQ ID NO: 15:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 26 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: unknown
- (ii) MOLECULE TYPE: peptide
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:
- | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Phe | Gln | Cys | Arg | Ile | Cys | Met | Arg | Asn | Phe |
| | | | | 5 | | | | | 10 |
| Ser | Gln | Ala | Gln | Thr | Leu | Gln | Arg | His | Leu |
| | | | | 15 | | | | | 20 |
| Lys | His | Thr | Gly | Glu | Lys | | | | |
| | | | | 25 | | | | | |

(2) INFORMATION FOR SEQ ID NO: 16:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 26 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: unknown
- (ii) MOLECULE TYPE: peptide
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:
- | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Phe | Gln | Cys | Arg | Ile | Cys | Met | Arg | Asn | Phe |
| | | | | 5 | | | | | 10 |
| Ser | Gln | Ala | Ala | Thr | Leu | Gln | Arg | His | Leu |
| | | | | 15 | | | | | 20 |
| Lys | His | Thr | Gly | Glu | Lys | | | | |
| | | | | 25 | | | | | |

(2) INFORMATION FOR SEQ ID NO: 17:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 26 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: unknown
- (ii) MOLECULE TYPE: peptide
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:
- | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Phe | Gln | Cys | Arg | Ile | Cys | Met | Arg | Asn | Phe |
| | | | | 5 | | | | | 10 |

Ser Gln Ala Gln Asp Leu Gln Arg His Leu
 15 20

Lys His Thr Gly Glu Lys
 25

(2) INFORMATION FOR SEQ ID NO: 18:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 89 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS:

(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

Met Ala Glu Glu Lys Pro Phe Gln Cys Arg
 5 10

Ile Cys Met Arg Asn Phe Ser Asp Arg Ser
 15 20

Ser Leu Thr Arg His Thr Arg Thr His Thr
 25 30

Gly Glu Lys Pro Phe Gln Cys Arg Ile Cys
 35 40

Met Arg Asn Phe Ser Asp Arg Ser His Leu
 45 50

Thr Arg His Thr Arg Thr His Thr Gly Glu
 55 60

Lys Pro Phe Gln Cys Arg Ile Cys Met Arg
 65 70

Asn Phe Ser Asp Arg Ser Asn Leu Thr Arg
 75 80

His Thr Arg Thr His Thr Gly Glu Lys
 85